

ON ROBUSTNESS OF THE SHIRYAEV–ROBERTS CHANGE-POINT DETECTION PROCEDURE UNDER PARAMETER MISSPECIFICATION IN THE POST-CHANGE DISTRIBUTION

Wenyu Du, Aleksey S. Polunchenko*, Grigory Sokolov

*Department of Mathematical Sciences, State University of New York (SUNY) at Binghamton
Binghamton, NY 13902–6000, USA*

Abstract

The gist of the quickest change-point detection problem is to detect the presence of a change in the statistical behavior of a series of sequentially made observations, and do so in an optimal detection-speed-vs.-“false-positive”-risk manner. When optimality is understood either in the generalized Bayesian sense or as defined in Shiryaev’s multi-cyclic setup, the so-called Shiryaev–Roberts (SR) detection procedure is known to be the “best one can do”, provided, however, that the observations’ pre- and post-change distributions are both fully specified. We consider a more realistic setup, viz. one where the post-change distribution is assumed known only up to a parameter, so that the latter may be “misspecified”. The question of interest is the sensitivity (or robustness) of the otherwise “best” SR procedure with respect to a possible misspecification of the post-change distribution parameter. To answer this question, we provide a case study where, in a specific Gaussian scenario, we allow the SR procedure to be “out of tune” in the way of the post-change distribution parameter, and numerically assess the effect of the “mistuning” on Shiryaev’s (multi-cyclic) Stationary Average Detection Delay delivered by the SR procedure. The comprehensive *quantitative* robustness characterization of the SR procedure obtained in the study can be used to develop the respective theory as well as to provide a rational for practical design of the SR procedure. The overall *qualitative* conclusion of the study is an expected one: the SR procedure is less (more) robust for less (more) contrast changes and for lower (higher) levels of the false alarm risk.

Keywords:

Sequential analysis, Shiryaev–Roberts procedure, Quality control, Quickest change-point detection.

*Address correspondence to A.S. Polunchenko, Department of Mathematical Sciences, State University of New York (SUNY) at Binghamton, Binghamton, NY 13902–6000, USA; Tel: +1 (607) 777-6906; Fax: +1 (607) 777-2450; Email: aleksey@binghamton.edu

Email addresses: wdu@binghamton.edu (Wenyu Du), aleksey@binghamton.edu (Aleksey S. Polunchenko), gsokolov@binghamton.edu (Grigory Sokolov)

URL: <http://www.math.binghamton.edu/grads/wdu> (Wenyu Du),
<http://www.math.binghamton.edu/aleksey> (Aleksey S. Polunchenko),
<http://www.math.binghamton.edu/gsokolov> (Grigory Sokolov)

1. Introduction

Sequential (quickest) change-point detection is concerned with the development and evaluation of statistical procedures for rapid and reliable “on-the-go” detection of unanticipated changes that may occur in the characteristics of a “live” process. Specifically, the process is assumed to be continuously monitored through sequentially made observations (e.g., measurements), and should at any point in time their behavior start to appear as though the process may have (been) statistically changed, the aim is to conclude so within the smallest number of observations possible, subject to a tolerable level of the false positive risk. As soon as such a conclusion is reached, an “alarm” is flagged, and an appropriate response action is taken (e.g., an investigation is initiated as to the possible cause of the alarm). For a thorough treatment of the subject’s theory as well as for examples of applications, see, e.g., Shiryaev (1978), Basseville and Nikiforov (1993), Poor and Hadjiliadis (2009), Veeravalli and Banerjee (2013), or (Tartakovsky et al., 2014, Parts II and III), and the references therein.

A sequential change-point detection procedure is identified with a stopping time, T , that is functionally dependent on the observations, $\{X_n\}_{n \geq 1}$; the semantics of T is that it constitutes a rule whereby one is to stop and declare that the statistical profile of the process under surveillance may have (been) changed. A “good” (i.e., optimal or nearly optimal) detection procedure is one that minimizes (or nearly minimizes) the desired detection delay penalty, subject to a constraint on the false alarm risk. For an overview of the major optimality criteria, see, e.g., Tartakovsky and Moustakides (2010), Polunchenko and Tartakovsky (2012), Veeravalli and Banerjee (2013), Polunchenko et al. (2013), or (Tartakovsky et al., 2014, Part II).

The basic quickest change-point detection problem assumes (a) that the observations, $\{X_n\}_{n \geq 1}$, are independent throughout the entire period of surveillance; (b) that the observations’ common pre-change probability density function (pdf) $f(x)$ is completely known; and (c) that so is the observations’ common post-change pdf $g(x) \neq f(x)$. This version of the problem is well-understood and has been solved (either exactly or asymptotically) under a variety of criteria. For a survey of the corresponding state-of-the-art, see, e.g., Tartakovsky and Moustakides (2010), Polunchenko and Tartakovsky (2012), Veeravalli and Banerjee (2013), Polunchenko et al. (2013), or (Tartakovsky et al., 2014, Part II).

This work’s focus is on a more realistic setup, namely, one where the observations’ post-change pdf, $g(x)$, is assumed known only up to a parameter; assumptions (a) and (b) above are retained. While being a more practical assumption, limited (parametric) knowledge of $g(x)$ makes the problem more difficult. In particular, the uncertainty in the post-change distribution parameter opens the door for a possible misspecification thereof. Consequently, when this parameter is set incorrectly (for whatever reason), the performance of any otherwise optimal detection procedure will naturally degrade. The aim of this work is to quantify the severity of this performance degradation as a function of the magnitude of the post-change distribution parameter misspecification and under various levels of the false alarm risk. To make it more clear, our intent in this paper is not to propose a way to deal with the parametric uncertainty in the post-change distribution, but merely to provide a *quantitative nonasymptotic* answer to the practically important question of robustness (or sensitivity) of a given detection procedure with respect to possible errors in the value of the post-change distribution parameter.

More concretely, our goal is to examine the robustness question specifically for the so-called Shiryaev–Roberts (SR) procedure, proposed (following quasi-Bayesian considerations) by Shiryaev (1961, 1963) and (independently) by Roberts (1966). The term “Shiryaev–Roberts” appears to have

been “coined” by Pollak (1985). For a brief account of the SR procedure’s history, see, e.g., Pollak (2009). The interest in the SR procedure is due to two reasons. First, the SR procedure is exactly optimal in Shiryaev’s (1961; 1963) multi-cyclic sense. Second, since Shiryaev’s multi-cyclic version of the change-point detection problem is equivalent to the latter’s generalized Bayesian setup, the SR procedure is exactly optimal in the generalized Bayesian sense as well. These results were first obtained by Shiryaev (1961, 1963) for a (continuous-time) Brownian motion drift-shift scenario; see also, e.g., Shiryaev (2002); Feinberg and Shiryaev (2006). For the discrete-time case, these results were recently established by Pollak and Tartakovsky (2008, 2009) and by Shiryaev and Zryumov (2010). It is noteworthy that neither the celebrated Cumulative Sum (CUSUM) “inspection scheme” proposed by Page (1954) nor the popular Exponentially Weighted Moving Average (EWMA) chart introduced by Roberts (1959) possesses such strong optimality properties. For comparative performance analyses of the SR procedure against the CUSUM scheme and the EWMA chart, see, e.g., Knoth (2006), Mahmoud et al. (2008), Moustakides et al. (2009), Tartakovsky et al. (2009), and Polunchenko et al. (2015). However, albeit strong, these optimality properties of the SR procedure hinge on the assumption of complete knowledge of the observations’ pre- and post-change distributions, whose pdf’s we agreed to denote by $f(x)$ and by $g(x)$, respectively. When the latter involves a parameter that the SR procedure has an incorrect value for, the performance of the SR procedure will no longer be “best” (optimal). Furthermore, the larger the error in the parameter value, the bigger the respective performance loss. It is to study this “cause-and-effect” relationship *quantitatively* that is the objective of this work, and the focus is entirely on the SR procedure.

Specifically, we offer a case study where, in a particular Gaussian scenario, we intentionally set up the SR procedure so that it “anticipates” the observations’ pdf to change from $f(x)$ pre-change to $g(x; \tilde{\theta})$ post-change, while the actual change is from $f(x)$ pre-change to $g(x; \theta)$ post-change, where $\tilde{\theta} \neq \theta$, and then assess the effect of the “mistuning” on the procedure’s performance. The performance characteristics of interest are the usual minimax Average Run Length (ARL) to false alarm proposed by Lorden (1971) (although see also Page 1954) and Shiryaev’s (1961; 1963) multi-cyclic Stationary Average Detection Delay (STADD). We compute both metrics numerically, with the aid of an integral-equations-based numerical method that is a hybrid between one proposed and used earlier by Moustakides et al. (2009, 2011) and a recent refinement thereof offered and stress-tested by Polunchenko et al. (2014b,a). The latter method was designed specifically for the SR procedure, and is more accurate when it comes to evaluation of the ARL to false alarm. The former method is used to compute the STADD, and even though it was originally designed only for the case when $\tilde{\theta} = \theta$, it extends trivially to the case when $\tilde{\theta} \neq \theta$. The study provides an exhaustive quantitative characterization of the SR procedure’s robustness. Qualitatively, the overall conclusion of the study is essentially what one would expect it to be: the less (more) contrast the change and the higher (lower) the level of the ARL to false alarm, the less (more) robust the SR procedure devised to detect it. To the best of our knowledge, when the observations are drawn at discrete points in time, no such study has previously been undertaken, even though its significance for applications is apparent. However, for similar studies in the continuous-time case, see, e.g., Pollak and Siegmund (1985) and Srivastava and Wu (1993).

The rest of the paper is organized as follows. Firstly, in Section 2 we formally state the problem, describe the SR procedure, briefly review its properties, and comment on how we intend to evaluate its performance. The case study itself is carried out next, in Section 3, which is the core section of the

paper. Lastly, in Section 4 we draw conclusions.

2. The problem and preliminary background

Since the version of the quickest change-point detection problem that is considered in this work is a build-up to the problem’s basic “minimax’ish” setup, it is convenient to briefly describe the latter first, i.e., start with the standard assumption of complete knowledge of the observations’ pre- and post-change distributions. Let the pdf’s of these distributions be $f(x)$ and $g(x)$, respectively; $g(x) \neq f(x)$. Define the change-point, $0 \leq \nu \leq \infty$, as the unknown (but not random) serial index of the final pre-change observation; note that it can potentially be infinite. That is, as illustrated in Figure 1, the pdf of X_n is $f(x)$ for $1 \leq n \leq \nu$, and $g(x)$ for $n \geq \nu + 1$. The notation $\nu = 0$ is to be understood as the case

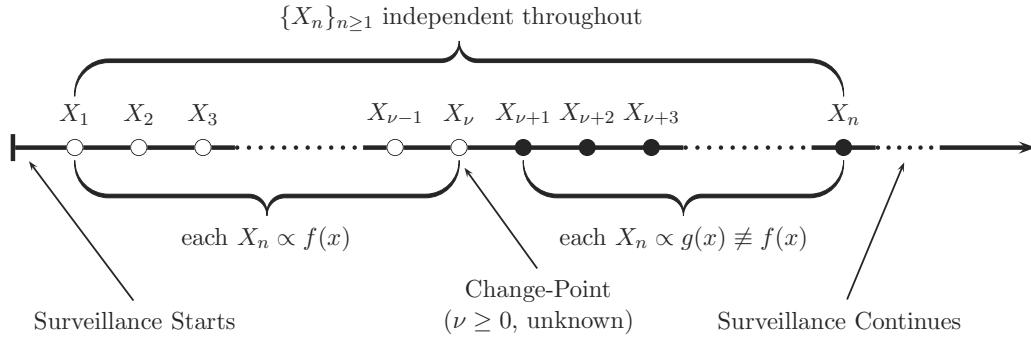


Figure 1: Basic “minimax-ish” setup of the quickest change-point detection problem.

when the pdf of X_n is $g(x)$ for all $n \geq 1$, i.e., the distributional pattern of the data, $\{X_n\}_{n \geq 1}$, is affected by change *ab initio*. Similarly, the notation $\nu = \infty$ is to mean that the pdf of X_n is $f(x)$ for all $n \geq 1$, i.e., the distributional pattern of the data, $\{X_n\}_{n \geq 1}$, never changes.

Let $\mathbb{P}_k(\mathbb{E}_k)$ be the probability measure (expectation) given that the change-point ν is at time moment (epoch) k , i.e., $\nu = k$, where $0 \leq k \leq \infty$. Particularly, $\mathbb{P}_\infty(\mathbb{E}_\infty)$ is the probability measure (expectation) when the observations always come from the pdf $f(x)$, i.e., $\nu = \infty$ so that there is never a change. Likewise, $\mathbb{P}_0(\mathbb{E}_0)$ is the probability measure (expectation) when the observations’ distribution is $g(x)$ from the start (i.e., $\nu = 0$).

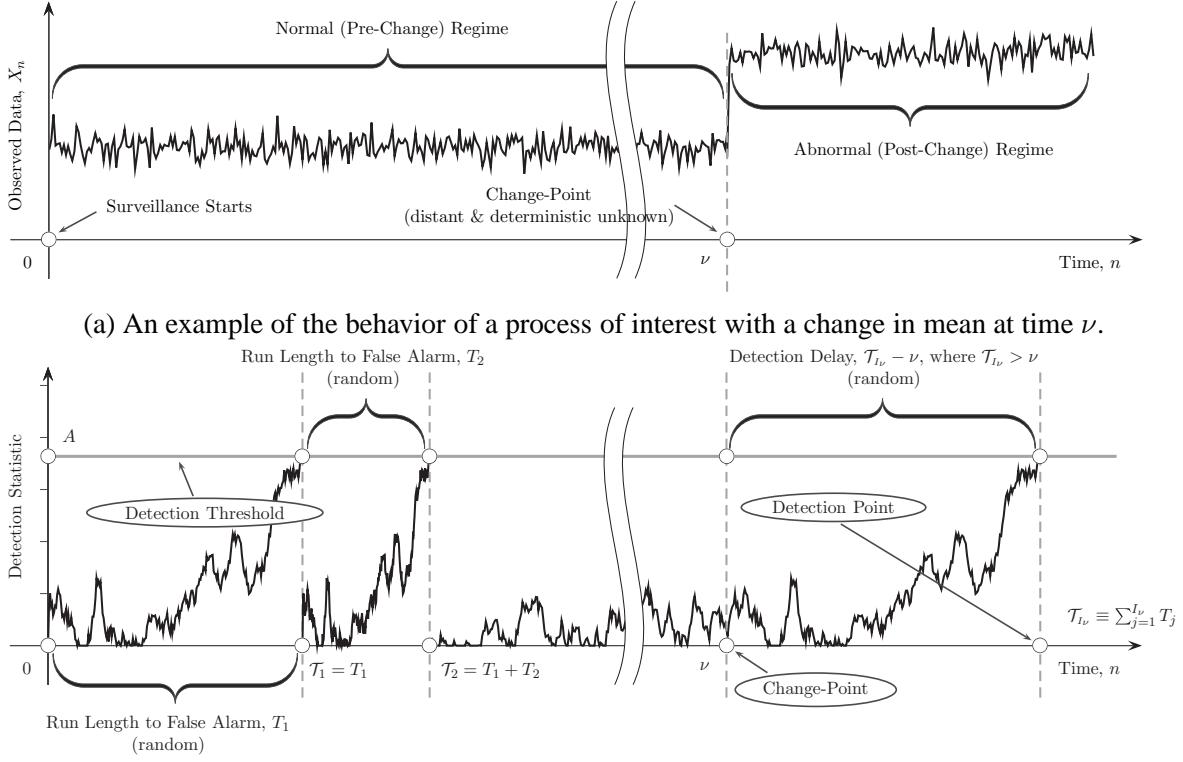
From now on T will denote the stopping time associated with a generic detection procedure.

Given this “minimax-ish” context, the standard way to gauge the false alarm risk is through Lorden’s (1971) ARL to false alarm (see also Page 1954). The ARL to false alarm is defined as $\text{ARL}(T) \triangleq \mathbb{E}_\infty[T]$ and captures the average number of observations the procedure (given by stopping time T) samples under the pre-change regime before stopping (i.e., before triggering a false alarm). The reciprocal of $\text{ARL}(T)$ can be interpreted (roughly) as the frequency of false alarms. Hence, the false alarm risk turns out to be inversely proportional to $\text{ARL}(T)$: the higher (lower) the latter, the lower (higher) the former.

To introduce the multi-cyclic change-point detection problem, let

$$\Delta(\gamma) \triangleq \left\{ T: \text{ARL}(T) \geq \gamma \right\}, \text{ where } \gamma > 1,$$

denote the class of procedures (stopping times), T , whose the ARL to false alarm is at least $\gamma > 1$, a pre-selected tolerance level. Suppose now that it is of utmost importance to detect the change as quickly as possible, even at the expense of raising many false alarms (using a repeated application of the same procedure) before the change occurs. Put otherwise, in exchange for the assurance that the change will be detected with maximal speed, one agrees to go through a “flurry” of false alarms along the way (the false alarms are ensued from repeatedly applying the same procedure, starting from scratch after each false alarm). This scenario is shown in Figure 2.



(b) Typical behavior of the detection statistic in the multi-cyclic mode.

Figure 2: Multi-cyclic change-point detection in a stationary regime.

Formally, let T_1, T_2, \dots be sequential independent repetitions of the same stopping time, T , and let $\mathcal{T}_j \triangleq T_1 + T_2 + \dots + T_j$, $j \geq 1$, be the time of the j -th alarm. Define $I_\nu \triangleq \min\{j \geq 1: \mathcal{T}_j > \nu\}$ so that \mathcal{T}_{I_ν} is the time of detection of a true change that occurs at time moment ν after $I_\nu - 1$ false alarms had been raised. One can then view the difference $\mathcal{T}_{I_\nu} - \nu (\geq 0)$ as the detection delay. Let

$$\text{STADD}(T) \triangleq \lim_{\nu \rightarrow \infty} \mathbb{E}_\nu [\mathcal{T}_{I_\nu} - \nu]$$

be the limiting value of the Average Detection Delay (ADD) referred to as the *Stationary ADD* (STADD); this metric was introduced by Shiryaev (1961, 1963). The multi-cyclic change-point detection problem

then is:

$$\text{to find } T_{\text{opt}} = \arg \inf_{T \in \Delta(\gamma)} \text{STADD}(T) \text{ for any given } \gamma > 1; \quad (1)$$

cf. Shiryaev (1961, 1963).

We note that, in this setup, $\text{ARL}(T)$ is exactly the average distance between successive false alarms. Therefore, $1/\text{ARL}(T)$ can be thought as the frequency of false alarms. Since higher (lower) frequency of false alarms clearly corresponds to higher (lower) false alarm risk, lower (higher) levels of $\text{ARL}(T)$ correspond to higher (lower) levels of the false alarm risk.

As can be gathered from the description, the “intrinsic assumption” of the multi-cyclic change-point detection problem is that the process under surveillance is not expected to be affected by change “for a while”, i.e., the change-point, ν , is large. Scenarios where this is a reasonable assumption may be encountered, e.g., in cybersecurity (see, e.g., Polunchenko et al. 2012 or Tartakovsky et al. 2013) and in the economic design of quality control charts (see, e.g., Duncan 1956; Montgomery 1980; Lorenzen and Vance 1986; Ho and Case 1994).

The multi-cyclic change-point detection problem (1) has been solved by Pollak and Tartakovsky (2008, 2009) and by Shiryaev and Zryumov (2010) who showed that the solution is the Shiryaev–Roberts (SR) procedure, due to Shiryaev (1961, 1963) and Roberts (1966). We reiterate that neither the CUSUM scheme nor the EWMA chart possesses such a strong optimality property. As a matter of fact, the comparative performance analyses carried out by Moustakides et al. (2009); Tartakovsky et al. (2009); Polunchenko et al. (2015) confirmed experimentally that both the CUSUM scheme and the EWMA chart are inferior to (i.e., are outperformed by) the SR procedure in the multi-cyclic sense. For similar analyses carried out in continuous time, see, e.g., Pollak and Siegmund (1985) and Srivastava and Wu (1993).

To introduce the SR procedure, let us first point out that, by and large, in all of statistics, there are two principally different ways to extract information from data to make inference. One of these approaches is based on maximization of a certain functional of the data. The other approach is based on aggregation of information in a sum-like manner. This is illustrated in Figure 3.

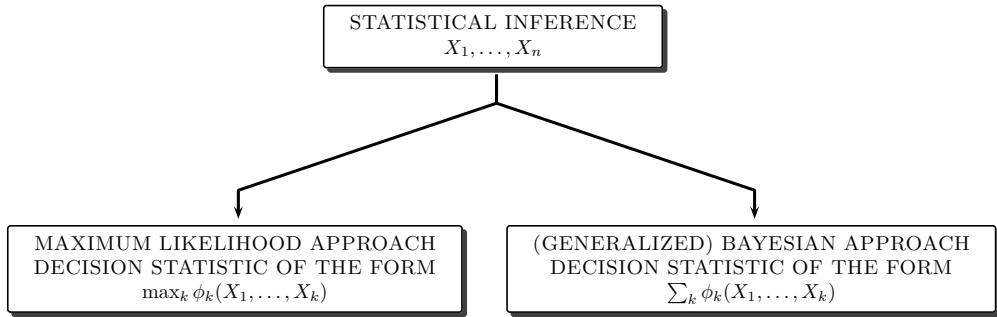


Figure 3: Two principally different approaches to statistical inference.

In change-point detection, the first approach is manifested in the CUSUM scheme, which uses the maximum-likelihood principle to “sense” whether the process under surveillance has gone “out-of-

control”. The EWMA chart is an example of a detection procedure whose decision statistic is based on aggregation of information in a Bayesian-like manner. The SR procedure is also of the latter type. It is formally defined as the stopping time

$$\mathcal{S}_A \triangleq \inf \{n \geq 1: R_n \geq A\}, \text{ such that } \inf \{\emptyset\} = \infty, \quad (2)$$

where $A > 0$ is a detection threshold (control limit) used to control the level of the ARL to false alarm, and

$$R_n \triangleq \sum_{k=1}^n \prod_{i=k}^n \Lambda_i, \quad n \geq 1, \quad (3)$$

is the SR detection statistic; here and onward $\Lambda_i \triangleq g(X_i)/f(X_i)$ denotes the “instantaneous” likelihood ratio (LR) for the i -th data point X_i . Note the recursion

$$R_{n+1} = (1 + R_n) \Lambda_{n+1} \text{ for } n = 0, 1, \dots \text{ with } R_0 = 0. \quad (4)$$

An important property of the SR statistic $\{R_n\}_{n \geq 0}$ is that the sequence $\{R_n - n\}_{n \geq 0}$ is a zero-mean \mathbb{P}_∞ -martingale, i.e., $\mathbb{E}_\infty[R_n - n] = 0$ for any $n \geq 0$. From this and Doob’s optional stopping (sampling) theorem (see, e.g., Shiryaev 1995, Chapter VII, Poor and Hadjiliadis 2009, Subsection 2.3.2 or Tartakovsky et al. 2014, Theorem 2.3.1, p. 31), one can conclude that $\mathbb{E}_\infty[R_{\mathcal{S}_A} - \mathcal{S}_A] = 0$ so that $\text{ARL}(\mathcal{S}_A) \triangleq \mathbb{E}_\infty[\mathcal{S}_A] = \mathbb{E}_\infty[R_{\mathcal{S}_A}] \geq A$. Hence, for any given $\gamma > 1$, if $A \geq \gamma$, then $\text{ARL}(\mathcal{S}_A) \geq \gamma$ and $\mathcal{S}_A \in \Delta(\gamma)$. As argued by Kenett and Pollak (1996), such a simple relation between $\text{ARL}(\mathcal{S}_A)$ and A endows the SR procedure with certain data-analytic advantages over the CUSUM scheme and the EWMA chart. Furthermore, note that all this is valid irrespective of the particular $f(x)$ and $g(x)$.

A more accurate connection between $\text{ARL}(\mathcal{S}_A)$ and A has been established by Pollak (1987) who showed that $\text{ARL}(\mathcal{S}_A) = (A/\zeta)[1+o(1)]$ as $A \rightarrow \infty$, i.e., $\text{ARL}(\mathcal{S}_A) \approx A/\zeta$ for sufficiently large A . To define ζ , let $S_n \triangleq \sum_{i=1}^n \log \Lambda_i$, $n \geq 1$, and let $\tau_a \triangleq \inf\{n \geq 1: S_n \geq a\}$, $a > 0$ (again, with the understanding that $\inf\{\emptyset\} = \infty$). Then $\kappa_a \triangleq S_{\tau_a} - a$ (≥ 0) is the so-called “overshoot” (excess over the level $a > 0$ at stopping), and $\zeta \triangleq \lim_{a \rightarrow \infty} \mathbb{E}_0[e^{-\kappa_a}]$, and is referred to as the “limiting average exponential overshoot”. In general, ζ is between 0 and 1, and the evaluation of this model-dependent constant falls within the scope of nonlinear renewal theory; see, e.g., Woodroffe (1982), (Veeravalli and Banerjee, 2013, Section II.C) or (Tartakovsky et al., 2014, Section 2.6).

More importantly, as was mentioned earlier, according to the result of Pollak and Tartakovsky (2008, 2009) and also to that of Shiryaev and Zryumov (2010), the SR procedure is *exactly* STADD(T)-optimal. That is, formally: $\text{STADD}(\mathcal{S}_{A_\gamma}) = \inf_{T \in \Delta(\gamma)} \text{STADD}(T)$ for every $\gamma > 1$, where $A_\gamma > 0$ is the solution of the equation $\text{ARL}(\mathcal{S}_{A_\gamma}) = \gamma > 1$. However, and we mentioned that already as well, this result ceases to hold when the SR statistic is “out of tune” in the way of either $f(x)$ or $g(x)$. We are interested in the case when $g(x)$ is given parametrically as $g(x; \theta)$, and the value of θ that is used by the SR statistic is $\tilde{\theta} \neq \theta$. To reflect that in this case $\text{STADD}(\mathcal{S}_{A_\gamma})$ is a function of both $\tilde{\theta}$ and θ , we will use the notation $\text{STADD}_{\tilde{\theta}, \theta}(\mathcal{S}_{A_\gamma})$. Then clearly $\text{STADD}_{\tilde{\theta}, \theta}(\mathcal{S}_{A_\gamma}) \geq \text{STADD}_{\theta, \theta}(\mathcal{S}_{A_\gamma})$ for all $\tilde{\theta} \neq \theta$. The aim of this work is to assess the sensitivity of $\text{STADD}_{\tilde{\theta}, \theta}(\mathcal{S}_{A_\gamma})$ with respect to the magnitude of

the difference between $\tilde{\theta}$ and θ and under various levels of the ARL to false alarm $\text{ARL}(\mathcal{S}_{A_\gamma}) = \gamma > 1$.

It is also noteworthy that the SR procedure (2)–(4) is optimal in the generalized Bayesian sense as well. Specifically, the generalized Bayesian change-point detection problem is: to find $T_{\text{opt}} = \arg \inf_{T \in \Delta(\gamma)} \text{RIADD}(T)$ for any given $\gamma > 1$, where $\text{RIADD}(T)$ is the *Relative Integral ADD* defined as

$$\text{RIADD}(T) \triangleq \frac{1}{\text{ARL}(T)} \sum_{\nu=0}^{\infty} \mathbb{E}_{\nu}[\max\{0, T - \nu\}]. \quad (5)$$

This version of the quickest change-point detection problem is a limiting case of the classical Bayesian change-point detection problem considered and solved by Shiryaev (1961, 1963). Specifically, the limit is with respect to the prior distribution of the change-point, ν , which is assumed to be improper uniform, so that each time instance becomes equally likely to be the change-point.

The exact generalized Bayesian optimality the SR procedure can be formally stated as follows: $\text{RIADD}(\mathcal{S}_{A_\gamma}) = \inf_{T \in \Delta(\gamma)} \text{RIADD}(T)$ for every $\gamma > 1$, where $A_\gamma > 0$ is again the solution of the equation $\text{ARL}(\mathcal{S}_{A_\gamma}) = \gamma > 1$. This result is also due to Pollak and Tartakovsky (2008, 2009) and Shiryaev and Zryumov (2010), and the proof is based on first showing that the generalized Bayesian version of the quickest change-point detection problem and its multi-cyclic setup (1) are equivalent in the sense that $\text{RIADD}(T) \equiv \text{STADD}(T)$ for any stopping time T , and then exploiting the fact that the SR procedure is exactly $\text{STADD}(T)$ -optimal.

The fact that $\text{RIADD}(T) \equiv \text{STADD}(T)$ for any stopping time T was used by Moustakides et al. (2011) to develop a numerical method to compute $\text{STADD}(\mathcal{S}_A)$, i.e., the STADD delivered by the SR procedure (2)–(4). Specifically, they observed that when T is based off of a Markovian detection statistic (note that the SR statistic $\{R_n\}$ is Markovian), the infinite sum appearing in the right-hand side of (5) is a convergent geometric series. This allowed Moustakides et al. (2011) to derive an integral (renewal) equation directly on the entire sum without any truncation, and then develop a numerical method to treat the integral equation; their method was recently extended by Polunchenko et al. (2014a). This is precisely how we intend to evaluate $\text{STADD}(\mathcal{S}_A)$ in our case study offered in the next section: $\text{STADD}(\mathcal{S}_A)$ will be computed numerically as $\text{RIADD}(\mathcal{S}_A)$ with the aid of the numerical methods of Moustakides et al. (2011) and Polunchenko et al. (2014a). We note that these numerical methods do not require any truncation of either the infinite sum appearing in the definition of $\text{RIADD}(T)$ or the limit appearing in the definition of $\text{STADD}(T)$. This fact makes both methods more accurate.

We would like to conclude the introduction of the SR procedure with a remark pertaining to a possible interpretation of the SR statistic, $\{R_n\}_{n \geq 0}$. Part of the reason why the SR procedure—in spite of its simplicity and strong optimality properties—has not yet received proper attention from engineers and applied scientists (in particular, from the quality control community), is that the SR statistic is problematic to chart. Specifically, the problem is that the decision-making mechanism behind the SR procedure is not as clearly understood as that behind, e.g., the CUSUM scheme or the EWMA chart, which have *de facto* been the industry “workhorse” for decades now. A simple intuitive explanation of how the SR statistic “works”, i.e., a meaningful “engineering” interpretation of the SR statistic, could help bridge this gap, and, in the long run, might also help “pave the way” for the SR procedure into the “engineering world”. An attempt to provide one such interpretation was previously made

by Kenett and Pollak (1996) who, in view of the quasi-Bayesian nature of the SR procedure, argued that the SR statistic, $\{R_n\}_{n \geq 0}$, may be regarded as a p -value.

The interpretation we have in mind for the SR statistic lends the latter a financial flavor, and involves an ordinary savings account and basic financial interest theory. Suppose first that one opens a savings account at a bank and immediately makes a deposit of \$1, so that at “time zero” (also referred to as “now” or “today”) the balance is \$1. If the account accrues interest at a rate of IR (decimals) per time period, then it is easy to see that the balance in the account “tomorrow” (i.e., in one time period from “today”) will be $\$1 \times (1 + \text{IR})$. This is schematically illustrated in Figure 4.

Now suppose that the time horizon is longer than just “one day”, and once the bank account is opened, one funds it not with one, but with a series of *periodic* single-dollar deposits. Moreover, suppose also that the interest rate *varies* from period to period. This is shown in Figure 5.

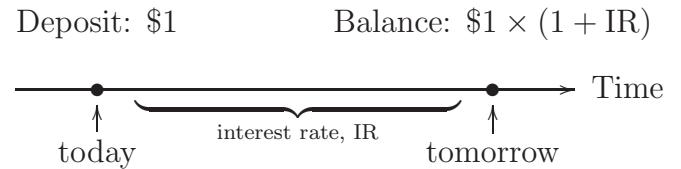


Figure 4: Financial analogy to interpret the SR statistic.

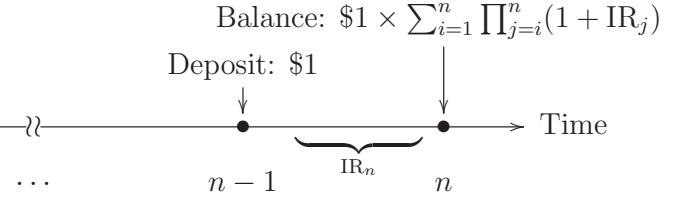
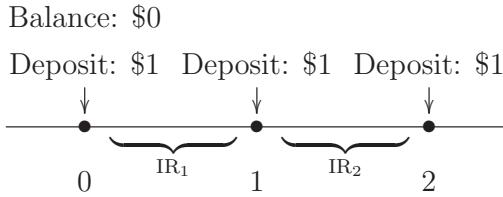


Figure 5: Financial analogy to interpret the SR statistic.

Then, at time n before the n -th deposit is made, the balance will be

$$B_n = \$1 \times \sum_{i=1}^n \prod_{j=i}^n (1 + \text{IR}_j), \quad n \geq 1,$$

and we hasten to note the similarity between this formula and that for the SR statistic (3): the two formulae are identical if the LR at the n -th epoch, Λ_n , is regarded as the n -th period interest factor, i.e., the quantity $(1 + \text{IR}_n)$. This suggests that the SR statistic, $\{R_n\}_{n \geq 0}$, can be thought of the balance in one’s savings account at epoch $n \geq 1$, assuming that the account is (a) funded through a series of periodic one-dollar deposits, and (b) credited (compound) interest every period. If the detection threshold, $A > 0$, is interpreted as a “target balance”, then the SR stopping time \mathcal{S}_A is the amount of time one is to wait before the account accumulates the target balance. Clearly, the higher the interest rate, the shorter the wait. This can now be used to explain how the SR statistic makes its decision as to the presence of a change in the observed process.

On the one hand, since $\mathbb{E}_\infty[\Lambda_n] = 1$ for all $n \geq 1$, this is equivalent to saying that IR_n is (on average) zero, and, therefore, if there is no change, then there is no interest, and the value of money does not change over time, so the balance in the account after n single-dollar deposits is simply the combined

amount thereof, which is $\$ n$. This is in perfect alignment with the well-known \mathbb{P}_∞ -martingale property of the SR procedure mentioned above, i.e., that $\mathbb{E}_\infty[R_n - n] = 0$ for any $n \geq 0$.

On the other hand, since $\mathbb{E}_0[\Lambda_n] > 1$ for all $n \geq 1$, this is the same as to say that IR_n is (on average) positive, and, therefore, the account is credited (compound) interest, so that the balance grows faster. Hence, if there is a change, the target balance of $\$A > 0$ is achieved sooner, resulting in a quicker detection of the change.

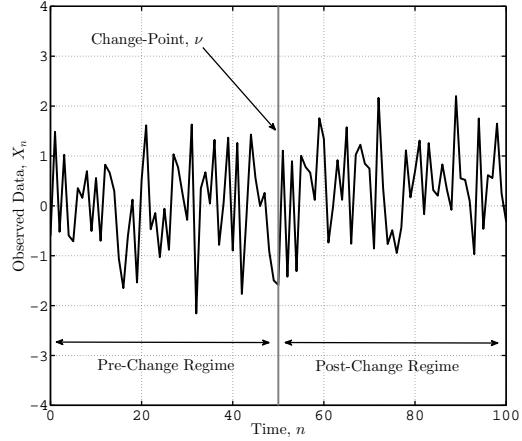
We would like to stress that this interpretation of the SR statistic is not to say that the SR procedure is “better” or “worse” than, e.g., the CUSUM scheme or the EWMA chart. The point is merely to help the “uninitiated” to intuitively understand why it is not unreasonable to at least expect the SR approach to work as a change-point detection tool to begin with. With this in mind, it is also instructional to mention that the decision-making process behind the CUSUM scheme is drastically different from that behind the SR procedure. Specifically, recall first that the CUSUM chart calls for stopping at $\mathcal{C}_A \triangleq \inf\{n \geq 1: W_n \geq A\}$, where $A > 0$ is again the detection threshold (control limit) and $\{W_n\}_{n \geq 0}$ is the CUSUM statistic defined as

$$W_n \triangleq \max\{0, W_{n-1} + \log \Lambda_n\} \text{ for } n = 0, 1, \dots \text{ with } W_0 = 0;$$

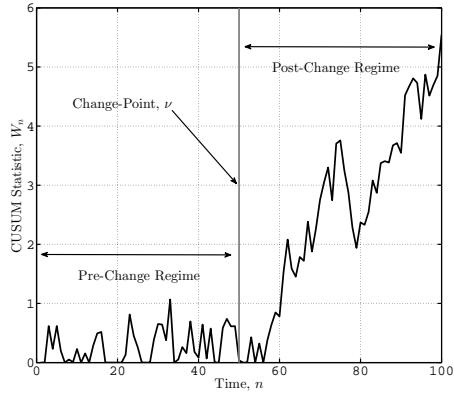
cf. Page (1954). Then, on the one hand, observe that because $\mathbb{E}_\infty[\log \Lambda_n] < 0$ for all $n \geq 1$, it follows that under the “no-change” regime the CUSUM statistic has a negative drift, which causes the CUSUM statistic to constantly “gravitate” toward zero. Since the latter is a reflective barrier for the CUSUM statistic, this endows the CUSUM scheme with a built-in mechanism to reset its statistic when, after a while of surveillance, the data have given no reason to believe their distribution changed. This internal resetting mechanism plays a major role in CUSUM’s minimax optimality first established by Moustakides (1986) and then also proved by Ritov (1990) using a different approach. By contrast, the SR procedure does not operate in this manner, and the SR statistic has no resetting feature to it. On the other hand, observe that $\mathbb{E}_\nu[\log \Lambda_n] > 0$ for all $n \geq 1$ and $0 \leq \nu < n$, and therefore, as soon as the observations are affected by change, the drift of the CUSUM statistic becomes positive, so that the CUSUM statistic starts to climb up toward the detection threshold, and eventually either hits it or surpasses it, signalling an alarm to indicate the change has (apparently) occurred. This difference in the way the CUSUM statistic behaves in the pre- and in the post-change regimes makes the CUSUM statistic very convenient to interpret when it is plotted against time: not only will the plot show that the change is likely to have occurred, but it will also provide a clue as to when it occurred. The SR statistic does not offer this kind of convenience.

To illustrate the above point, consider Figure 6 which gives an example of the typical behavior of the CUSUM statistic $\{W_n\}$ and that of the SR statistic $\{R_n\}$ for the problem of detecting a shift in the mean of a series of independent standard Gaussian observations $\{X_n\}$. Specifically, Figure 6a presents a sample trajectory of $\{X_n\}$ of 100 points, with the change occurring at epoch $\nu = 50$, i.e., right down the middle of the sample, and the magnitude of the shift is 0.5. Shown underneath Figure 6a are Figures 6b and 6c which depict the respective trajectories of the CUSUM statistic $\{W_n\}$ and of the SR statistic $\{R_n\}$.

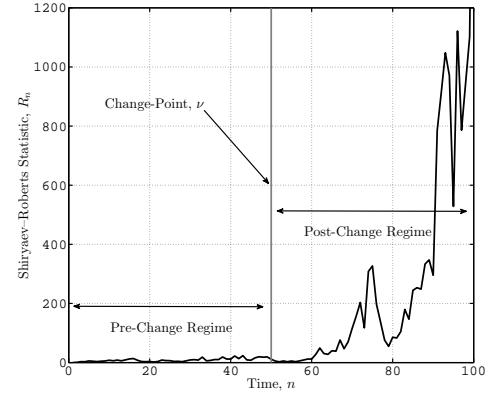
As can be seen, both the CUSUM statistic and the SR statistic change their behavior drastically as soon as the data series undergoes a change in the mean. Both statistics begin to climb up, and if there



(a) Observed data, $\{X_n\}$.



(b) CUSUM statistic, $\{W_n\}$.



(c) SR statistic, $\{R_n\}$.

Figure 6: Illustration of the behavior of the CUSUM statistic and that of the SR statistic for the same data.

were a detection threshold (control limit) imposed, one could declare the change as having occurred as soon as the respective statistic would reach or exceed the threshold. However, as was explained above, at the local level the two statistics are poles apart: they each “sense” the presence of the change in a completely different manner.

3. The case study

As the centerpiece of this work, this section is devoted to a case study, where, in a concrete scenario and subject to the constraint $ARL(\mathcal{S}_A) = \gamma$ for a given $\gamma > 1$, we quantify $STADD_{\tilde{\theta}, \theta}(\mathcal{S}_A)$ as a function of $\tilde{\theta}$ and θ , and, in particular, assess the sensitivity of $STADD_{\tilde{\theta}, \theta}(\mathcal{S}_A)$ with respect to the magnitude of the difference $\tilde{\theta} - \theta$. We would like to reiterate that to compute $ARL(\mathcal{S}_A)$ and $STADD_{\tilde{\theta}, \theta}(\mathcal{S}_A)$, we will rely on two numerical methods: one proposed by Moustakides et al. (2011) and its refinement due to Polunchenko et al. (2014b,a).

Specifically, suppose that the observations, $\{X_n\}_{n \geq 1}$, are independent and Gaussian-distributed, with mean zero pre-change and with mean $\theta \neq 0$ post-change, i.e., θ is the actual (true) value of the post-change mean; suppose also that the variance is 1 and does not change. Formally, the pre- and post-change pdf’s in this case are

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \text{ and } g(x; \theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x - \theta)^2}{2}\right\},$$

respectively, where $x \in \mathbb{R}$ and $\theta \neq 0$. When it is assumed that θ has no way to be misspecified, this Gaussian scenario is the standard “testbed” model widely used in the literature. We are interested in the case when θ *can* be misspecified (for whatever reason), and let $\tilde{\theta}$ denote the respective putative (or “anticipated”) value of θ . That is, it is $\tilde{\theta}$ (and not θ) that is to be used to construct the corresponding LR to then base the SR statistic off of.

For the Gaussian scenario under consideration, the corresponding “instantaneous” LR for the n -th data point, X_n , i.e., $\Lambda_n \triangleq g(X_n; \tilde{\theta})/f(X_n)$, can be seen to be

$$\Lambda_n = \exp\left\{\tilde{\theta}X_n - \frac{\tilde{\theta}^2}{2}\right\}, \quad n \geq 1,$$

and, therefore, for each $n \geq 1$, the LR’s distribution is log-normal and such that $\log \Lambda_n$ has mean $-\tilde{\theta}^2/2$ and variance $\tilde{\theta}^2$ under measure \mathbb{P}_∞ , and mean $\tilde{\theta}\theta - \tilde{\theta}^2/2$ ($\neq -\tilde{\theta}^2/2$) and variance $\tilde{\theta}^2$ under measure $\mathbb{P}_0(\cdot|\theta)$, where $\mathbb{P}_0(\cdot|\theta)$ denotes the probability measure under the assumption that the change is in effect from the start and that the actual distribution of the observations is $g(x; \theta)$. Specifically, let $P_\infty^\Lambda(t) \triangleq \mathbb{P}_\infty(\Lambda_1 \leq t)$, $t \geq 0$, and $P_0^\Lambda(t|\tilde{\theta}, \theta) \triangleq \mathbb{P}_0(\Lambda_1 \leq t|\theta)$, $t \geq 0$, be the cumulative distribution functions (cdf’s) of the LR in the pre- and in the post-change regimes, respectively. Then, we obtain:

$$P_\infty^\Lambda(t) = \begin{cases} \Phi\left(\text{sign}(\tilde{\theta})\left[\frac{1}{\tilde{\theta}} \log t + \frac{\tilde{\theta}}{2}\right]\right), & \text{for } t \geq 0; \\ 0, & \text{for } t < 0, \end{cases}$$

and

$$P_0^\Lambda(t|\tilde{\theta}, \theta) = \begin{cases} \Phi\left(\text{sign}(\tilde{\theta})\left[\frac{1}{\tilde{\theta}}\log t + \frac{\tilde{\theta}}{2} - \theta\right]\right), & \text{for } t \geq 0; \\ 0, & \text{for } t < 0, \end{cases}$$

where

$$\Phi(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad x \in \mathbb{R},$$

is the standard Gaussian cdf. We note that, as expected, $P_\infty^\Lambda(t)$ does not depend on θ . With $P_\infty^\Lambda(t)$ and $P_0^\Lambda(t|\tilde{\theta}, \theta)$ both expressed explicitly, the numerical method of Moustakides et al. (2011) and that of Polunchenko et al. (2014b,a) can be implemented. We implemented the methods in MATLAB, the popular software package for scientific simulation and computation developed by MathWorks, Inc. Both methods were set up so that the accuracy is on the order of a fraction of a percent.

We are now in a position to begin our case study. To that end, the obvious point of departure is the question: How does one choose the detection threshold, $A > 0$, so as to guarantee that $\text{ARL}(\mathcal{S}_A) = \gamma$ for a given $\gamma > 1$? To answer this question, recall that, as was discussed in Section 2, at least when $A > 0$ is sufficiently large, $\text{ARL}(\mathcal{S}_A) \approx A/\zeta$, where ζ is the limiting average exponential overshoot. Thus, if ζ were known (at least approximately), then $A \approx \zeta\gamma$ would provide (approximately) the sought value of the detection threshold needed to ensure $\text{ARL}(\mathcal{S}_A) \approx \gamma$ for a given $\gamma > 1$. For our Gaussian scenario, $\zeta = \zeta(\tilde{\theta})$ can be computed numerically from the well-known (exact) formula:

$$\zeta = \frac{2}{\tilde{\theta}^2} \exp\left\{-2 \sum_{k=1}^{\infty} \frac{1}{k} \Phi\left(-\frac{\tilde{\theta}}{2}\sqrt{k}\right)\right\},$$

where $\Phi(x)$ is the standard Gaussian cdf; see, e.g., (Woodroffe, 1982, Example 3.1, pp. 32–33) or (Tartakovsky et al., 2014, Section 3.1.5, p. 137). Since it is well known that

$$2\Phi(-t) \geq \sqrt{\frac{2}{\pi}} \frac{t}{1+t^2} e^{-\frac{t^2}{2}}, \quad \text{for all } t > 0,$$

it is easy to see that the infinite series appearing under the exponent in the above formula for ζ converges exponentially fast. Hence, to compute ζ accurately, it is “safe” to simply “chop off” the series at a reasonably distant term. We used the first 10^6 terms, which is substantially more than needed to compute ζ very accurately. We computed and tabulated ζ for $\tilde{\theta} = 0.1, 0.2, \dots, 1.0$. The obtained values (rounded to the first six decimal places) are reported in the second leftmost column in Table 1. These values coincide with those previously found by (Woodroffe, 1982, p. 33). With ζ known, the detection threshold, $A = A_\gamma$, needed to ensure $\text{ARL}(\mathcal{S}_{A_\gamma}) \approx \gamma$ for a given $\gamma > 1$ can be determined as $A_\gamma = \zeta\gamma$. However, note that the latter formula is an approximate one (it becomes exact only asymptotically, when $A \rightarrow \infty$), and if $A_\gamma = \zeta\gamma$ and is a *finite* number, then the actual proximity of $\text{ARL}(\mathcal{S}_{A_\gamma})$ to γ needs to be verified. To that end, for greater certainty, we used the numerical method of Polunchenko et al. (2014b) and computed numerically (with high accuracy) the actual level of the

ARL to false alarm for each value of the detection threshold. These numerically computed ARL to false alarm levels can be considered trustworthy. Table 1 reports the obtained results. Specifically, each column under the heading “Desired level of the Average Run Length (ARL) to false alarm” corresponds to a specific level of the ARL to false alarm $\gamma > 1$ picked between 10^2 and 10^4 . Each row corresponds to a specific value of $\tilde{\theta}$ (the value of θ in this case is irrelevant). The content of each cell is two numbers placed one on top of the other. The top number is the detection threshold computed as $A = \gamma\zeta$ for the corresponding γ and $\tilde{\theta}$. The bottom number (in parentheses) is the actual level of the ARL to false alarm (computed numerically using the numerical method of Polunchenko et al. 2014b) that corresponds to that particular detection threshold. The main conclusion that can be made from the results is that the approximation $\text{ARL}(\mathcal{S}_A) \approx A/\zeta$ is quite accurate, uniformly across all $\tilde{\theta}$ and all $\gamma > 1$.

We are now in a position to examine the sensitivity of $\text{STADD}_{\tilde{\theta},\theta}(\mathcal{S}_A)$ with respect to $\theta - \tilde{\theta}$, i.e., the size of the error in the post-change mean. Specifically, in view of the fact that $\text{STADD}_{\tilde{\theta},\theta}(\mathcal{S}_A)$ is the lowest when $\theta - \tilde{\theta} = 0$, it makes sense to use $\text{STADD}_{\theta,\theta}(\mathcal{S}_A)$ as a benchmark and measure the performance loss via the following relative efficiency (RE) metric:

$$\text{RE}_{\tilde{\theta},\theta}(\mathcal{S}_A) \triangleq \frac{\text{STADD}_{\tilde{\theta},\theta}(\mathcal{S}_A) - \text{STADD}_{\theta,\theta}(\mathcal{S}_A)}{\text{STADD}_{\theta,\theta}(\mathcal{S}_A)},$$

where it is understood that the constraint $\text{ARL}(\mathcal{S}_A) = \gamma$ is fulfilled for a given $\gamma > 1$. Due to the exact STADD-optimality of the SR procedure when $\tilde{\theta} = \theta$, it is clear that $\text{RE}_{\tilde{\theta},\theta}(\mathcal{S}_A) \geq 0$ in general, and that $\text{RE}_{\theta,\theta}(\mathcal{S}_A) = 0$ in particular. The actual robustness of the SR procedure would then be inversely proportional to $\text{RE}_{\tilde{\theta},\theta}(\mathcal{S}_A)$: smaller (higher) values of $\text{RE}_{\tilde{\theta},\theta}(\mathcal{S}_A)$ would correspond to greater (lower) robustness. We will consider three levels of the ARL to false alarm: $\gamma = 10^2$, which corresponds to high false alarm risk, $\gamma = 10^3$, which corresponds to moderate false alarm risk, and $\gamma = 10^4$, which corresponds to low false alarm risk.

Let us first consider the case when $\text{ARL}(\mathcal{S}_A) = \gamma = 10^2$. The values of the detection threshold for different values of $\tilde{\theta}$ are given in Table 1. Table 2 summarizes of the behavior of $\text{STADD}_{\tilde{\theta},\theta}(\mathcal{S}_A)$ and that of $\text{RE}_{\tilde{\theta},\theta}(\mathcal{S}_A)$ each regarded as a bivariate function of $\tilde{\theta}$ and θ . Specifically, the content of each cell in the table is a pair of numbers, one on top of the other: the top number is the actual $\text{STADD}_{\tilde{\theta},\theta}(\mathcal{S}_A)$ -value computed for the respective pair of $\tilde{\theta}$ and θ and rounded to two decimal places, and the bottom number (given in parentheses) is the respective $\text{RE}_{\tilde{\theta},\theta}(\mathcal{S}_A)$ -value given as a percentage. For example, the top number in cell positioned at the intersection of the first (topmost) row and the rightmost column is 9.86. This number is the value of $\text{STADD}_{\tilde{\theta},\theta}(\mathcal{S}_A)$ when $\tilde{\theta} = 0.1$ and $\theta = 1.0$ (and the ARL to false alarm is 10^2). The number underneath it is 80.68%, and it is the relative excess of the 9.86 over the ideal-case performance. The ideal case is when $\tilde{\theta} = \theta = 1.0$, and the respective value of $\text{STADD}_{\tilde{\theta},\theta}(\mathcal{S}_A)$ is 5.46 as can be inferred from the content of the cell positioned at the intersection of the bottommost row and the rightmost column. The 80.68 % is computed as $100 \% \times (9.86 - 5.46)/5.46$. This explains why all the percentages along the table’s diagonal are zero: it is the ideal case, i.e., when $\tilde{\theta} = \theta$, and in this case the SR procedure is exactly $\text{STADD}(T)$ -optimal.

As a complement to Table 2, Figure 7 provides a graphical representation of $\text{RE}_{\tilde{\theta},\theta}(\mathcal{S}_A)$ for various values of $\tilde{\theta}$ and θ by means of a grayscale “heat diagram”. The idea is to represent lower (higher) values of $\text{RE}_{\tilde{\theta},\theta}(\mathcal{S}_A)$ with fainter (stronger) shades of gray. For better perception, Figure 7 also contains

Table 1: Characterization of the detection threshold ($A > 0$) for selected values of the post-change mean ($\tilde{\theta} = \theta$) and ARL to false alarm (i.e., $\text{ARL}(\mathcal{S}_A) = \gamma > 1$).

	ζ	Desired level of Average Run Length (ARL) to false alarm (i.e., $\text{ARL}(\mathcal{S}_A) = \gamma > 1$)										
		1×10^2	2×10^2	3×10^2	4×10^2	5×10^2	6×10^2	7×10^2	8×10^2	9×10^2	1×10^3	
Post-change mean (putative=true)	0.1 0.943408	94.34 (100.28)	188.68 (200.28)	283.02 (300.28)	377.36 (400.28)	471.7 (500.28)	566.04 (600.28)	660.38 (700.28)	754.72 (800.27)	849.06 (900.27)	943.4 (1,000.27)	9,434.08 (10,000.28)
	0.2 0.890037	89.0 (100.31)	178.00 (200.31)	267.01 (300.32)	356.01 (400.31)	445.01 (500.31)	534.02 (600.31)	623.02 (700.31)	712.02 (800.31)	801.03 (900.31)	890.03 (1,000.31)	8,900.36 (10,000.3)
	0.3 0.839721	83.97 (100.35)	167.94 (200.35)	251.91 (300.35)	335.88 (400.35)	419.86 (500.35)	503.83 (600.35)	587.8 (700.35)	671.77 (800.35)	755.74 (900.34)	839.72 (1,000.35)	8,397.21 (10,000.36)
	0.4 0.792298	79.22 (100.39)	158.45 (200.39)	237.68 (300.39)	316.91 (400.39)	396.14 (500.39)	475.37 (600.39)	554.6 (700.39)	633.83 (800.39)	713.06 (900.39)	792.29 (1,000.39)	7,922.98 (10,000.4)
	0.5 0.747615	74.76 (100.45)	149.52 (200.44)	224.28 (300.44)	299.04 (400.44)	373.8 (500.44)	448.56 (600.44)	523.33 (700.45)	598.09 (800.44)	672.85 (900.44)	747.61 (1,000.44)	7,476.15 (10,000.45)
	0.6 0.705525	70.55 (100.5)	141.1 (200.5)	211.65 (300.5)	282.2 (400.5)	352.76 (500.5)	423.31 (600.5)	493.86 (700.5)	564.41 (800.5)	634.97 (900.5)	705.52 (1000.5)	7,055.25 (10,000.5)
	0.7 0.665887	66.58 (100.55)	133.17 (200.55)	199.76 (300.55)	266.35 (400.56)	332.94 (500.57)	399.53 (600.56)	466.12 (700.56)	532.7 (800.55)	599.29 (900.55)	665.88 (1,000.55)	6,658.87 (10,000.56)
	0.8 0.628566	62.85 (100.62)	125.71 (200.62)	188.56 (300.61)	251.42 (400.62)	314.28 (500.62)	377.13 (600.61)	439.99 (700.62)	502.85 (800.62)	565.7 (900.61)	628.56 (1,000.62)	6,285.66 (10,000.62)
	0.9 0.593435	59.34 (100.7)	118.68 (200.7)	178.03 (300.7)	237.37 (400.7)	296.71 (500.7)	356.06 (600.7)	415.4 (700.7)	474.74 (800.69)	534.09 (900.7)	593.43 (1,000.7)	5,934.35 (10,000.71)
	1.0 0.56037	56.03 (100.77)	112.07 (200.78)	168.11 (300.79)	224.14 (400.77)	280.18 (500.78)	336.22 (600.78)	392.25 (700.77)	448.29 (800.78)	504.33 (900.78)	560.37 (1,000.79)	5,603.7 (10,000.78)

Table 2: Characterization of $\text{STADD}_{\tilde{\theta}, \theta}(\mathcal{S}_A)$ and $\text{RE}_{\tilde{\theta}, \theta}(\mathcal{S}_A)$ as functions of the putative value $(\tilde{\theta})$ and of the true value (θ) of the post-change mean for the ARL to false alarm of 10^2 (i.e., $\text{ARL}(\mathcal{S}_A^r) = \gamma = 10^2$).

		True value of the post-change mean (θ)									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Putative value of the post-change mean $(\tilde{\theta})$	0.1	40.14 (0 %)	29.7 (5.18 %)	23.58 (14.29 %)	19.57 (24.2 %)	16.75 (34.18 %)	14.67 (44.0 %)	13.05 (53.58 %)	11.77 (62.89 %)	10.73 (71.92 %)	9.86 (80.68 %)
	0.2	41.93 (4.47 %)	28.24 (0 %)	21.01 (1.87 %)	16.67 (5.8 %)	13.81 (10.59 %)	11.79 (15.78 %)	10.3 (21.12 %)	9.14 (26.52 %)	8.23 (31.89 %)	7.49 (37.21 %)
	0.3	44.58 (11.06 %)	28.74 (1.78 %)	20.63 (0 %)	15.92 (1.03 %)	12.92 (3.4 %)	10.86 (6.6 %)	9.36 (10.17 %)	8.24 (13.99 %)	7.36 (17.93 %)	6.66 (21.93 %)
	0.4	47.1 (17.35 %)	29.74 (5.31 %)	20.84 (1.05 %)	15.76 (0 %)	12.57 (0.69 %)	10.43 (2.42 %)	8.91 (4.77 %)	7.77 (7.52 %)	6.9 (10.52 %)	6.21 (13.67 %)
	0.5	49.41 (23.09 %)	30.93 (9.52 %)	21.34 (3.47 %)	15.88 (0.74 %)	12.49 (0 %)	10.24 (0.52 %)	8.66 (1.85 %)	7.5 (3.73 %)	6.61 (5.96 %)	5.92 (8.43 %)
	0.6	51.51 (28.32 %)	32.2 (14.05 %)	22.01 (6.7 %)	16.17 (2.58 %)	12.56 (0.58 %)	10.18 (0 %)	8.53 (0.41 %)	7.34 (1.49 %)	6.43 (3.03 %)	5.73 (4.9 %)
	0.7	53.43 (33.12 %)	33.52 (18.72 %)	22.79 (10.46 %)	16.58 (5.2 %)	12.74 (2.06 %)	10.23 (0.47 %)	8.5 (0 %)	7.25 (0.33 %)	6.32 (1.23 %)	5.6 (2.54 %)
	0.8	55.21 (37.55 %)	34.85 (23.43 %)	23.64 (14.6 %)	17.08 (8.39 %)	13.02 (4.24 %)	10.36 (1.7 %)	8.53 (0.39 %)	7.23 (0 %)	6.26 (0.28 %)	5.52 (1.04 %)
	0.9	56.86 (41.65 %)	36.18 (28.13 %)	24.55 (19.01 %)	17.66 (12.04 %)	13.36 (6.98 %)	10.55 (3.56 %)	8.62 (1.44 %)	7.25 (0.34 %)	6.24 (0 %)	5.47 (0.23 %)
	1.0	58.39 (45.48 %)	37.5 (32.79 %)	25.49 (23.59 %)	18.29 (16.05 %)	13.76 (10.19 %)	10.79 (5.94 %)	8.76 (3.05 %)	7.32 (1.25 %)	6.26 (0.3 %)	5.46 (0 %)

contours corresponding to a few selected levels of $\text{RE}_{\tilde{\theta}, \theta}(\mathcal{S}_A)$. For instance, the curves labeled “5%” provide boundaries for the values of $\tilde{\theta}$ and θ for which $\text{RE}_{\tilde{\theta}, \theta}(\mathcal{S}_A) \leqslant 5\%$. One may note that more or less uniformly across the entire Table 2 (i.e., uniformly in $\tilde{\theta}$ and θ), if the error in the post-change mean is about $0.1 \div 0.2$ in absolute value, then the respective relative performance degradation is on the order of a few percentage points, which would be considered “tolerable” in most practical situations. Figure 7 confirms this. One may also note that the sensitivity is higher when θ and $\tilde{\theta}$ are small, and then it gradually decreases as θ and $\tilde{\theta}$ increase. This is especially apparent in Figure 7, where the “5 %”-level boundaries are narrower around the origin, and get further apart as one moves away from the origin. We also note that when $|\theta - \tilde{\theta}| \geqslant 0.5$, the performance loss may be hard to ignore in practice, as it is on the order of tens of percent.

Let us now consider the cases when $\gamma = 10^3$ and $\gamma = 10^4$. Tables 3 and 4 and Figures 8 and 9 summarize the corresponding results. By and large, the results are similar to the case when $\gamma = 10^2$. The difference is that the STADD numbers are higher for higher levels of the ARL to false alarm. This makes perfect sense, because the higher the ARL to false alarm level, the lower the false alarm risk, and the price to pay for being more risk-averse is a higher detection delay. More importantly, Tables 3 and 4 and Figures 8 and 9 suggest that, as the ARL to false alarm level increases, the SR procedure becomes more sensitive (or less robust) to the error in the post-change parameter. This conclusion can be drawn from the observation that, as the ARL to false alarm level increases, so do the $\text{RE}_{\tilde{\theta}, \theta}(\mathcal{S}_A)$

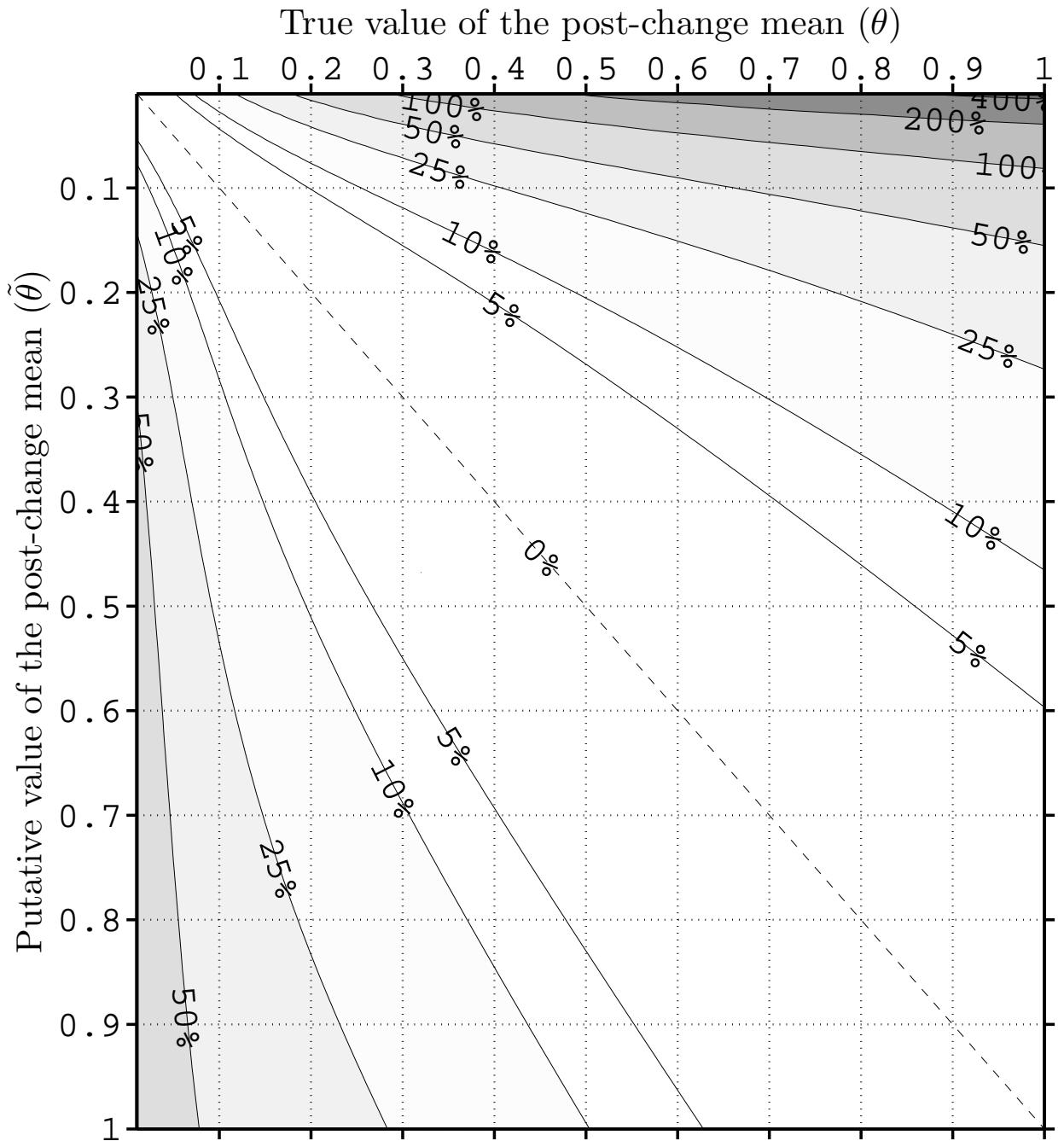


Figure 7: Characterization of $\text{RE}_{\tilde{\theta}, \theta}(S_A)$ as a function of the putative value ($\tilde{\theta}$) and of the true value (θ) of the post-change mean for the Gaussian scenario for the ARL to false alarm of 10^2 (i.e., $\text{ARL}(S_A^r) = \gamma = 10^2$).

numbers, uniformly across all $\tilde{\theta}$'s and θ 's. This effect of the ARL to false alarm level on the robustness of the SR procedure aligns well with one's intuition.

Table 3: Characterization of $\text{STADD}_{\tilde{\theta}, \theta}(\mathcal{S}_A)$ and $\text{RE}_{\tilde{\theta}, \theta}(\mathcal{S}_A)$ as functions of the putative value ($\tilde{\theta}$) and of the true value (θ) of the post-change mean for the ARL to false alarm of 10^3 (i.e., $\text{ARL}(\mathcal{S}_A^r) = \gamma = 10^3$).

		True value of the post-change mean (θ)									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Putative value of the post-change mean ($\tilde{\theta}$)	0.1	193.5 (0 %)	99.38 (6.92 %)	66.23 (18.96 %)	49.63 (32.0 %)	39.7 (45.13 %)	33.09 (58.05 %)	28.38 (70.7 %)	24.86 (83.03 %)	22.12 (95.04 %)	19.94 (106.75 %)
	0.2	206.82 (6.88 %)	92.94 (0 %)	57.48 (3.25 %)	41.31 (9.87 %)	32.2 (17.72 %)	26.38 (26.01 %)	22.35 (34.44 %)	19.4 (42.84 %)	17.15 (51.17 %)	15.37 (59.37 %)
	0.3	231.28 (19.52 %)	96.3 (3.61 %)	55.68 (0 %)	38.36 (2.03 %)	29.12 (6.47 %)	23.45 (11.99 %)	19.63 (18.02 %)	16.88 (24.29 %)	14.82 (30.64 %)	13.21 (37.01 %)
	0.4	258.75 (33.72 %)	104.22 (12.13 %)	56.96 (2.3 %)	37.6 (0 %)	27.74 (1.41 %)	21.91 (4.64 %)	18.09 (8.81 %)	15.41 (13.47 %)	13.43 (18.4 %)	11.91 (23.46 %)
	0.5	286.46 (48.04 %)	114.89 (23.62 %)	60.25 (8.21 %)	38.2 (1.6 %)	27.35 (0 %)	21.16 (1.05 %)	17.21 (3.52 %)	14.51 (6.8 %)	12.54 (10.54 %)	11.05 (14.55 %)
	0.6	313.25 (61.88 %)	127.31 (36.98 %)	65.03 (16.81 %)	39.82 (5.9 %)	27.68 (1.18 %)	20.94 (0 %)	16.76 (0.81 %)	13.96 (2.78 %)	11.96 (5.43 %)	10.46 (8.52 %)
	0.7	338.67 (75.02 %)	140.82 (51.52 %)	70.98 (27.48 %)	42.26 (12.4 %)	28.57 (4.44 %)	21.13 (0.91 %)	16.63 (0 %)	13.67 (0.65 %)	11.6 (2.25 %)	10.07 (4.46 %)
	0.8	362.57 (87.37 %)	154.98 (66.75 %)	77.84 (39.8 %)	45.4 (20.77 %)	29.95 (9.48 %)	21.66 (3.46 %)	16.75 (0.73 %)	13.58 (0 %)	11.4 (0.53 %)	9.82 (1.87 %)
	0.9	384.98 (98.95 %)	169.47 (82.34 %)	85.43 (53.44 %)	49.16 (30.75 %)	31.76 (16.11 %)	22.5 (7.47 %)	17.09 (2.78 %)	13.66 (0.59 %)	11.34 (0 %)	9.69 (0.44 %)
	1.0	405.96 (109.79 %)	184.1 (98.08 %)	93.6 (68.12 %)	53.45 (42.17 %)	33.97 (24.18 %)	23.62 (12.81 %)	17.63 (6.04 %)	13.89 (2.28 %)	11.4 (0.49 %)	9.64 (0 %)

4. Conclusion

This work sought to examine what happens to the performance of the Shiryaev–Roberts (SR) procedure when the latter is set up to detect the “wrong” change. Specifically, considered within the context of the basic quickest change-point detection problem, the SR procedure was intentionally allowed to be “out of tune” in the way of the actual value of a parameter in the observations’ post-change distribution (e.g., as a result of the latter known only up to that parameter), and the question was to understand how sensitive Shiryaev’s (1961; 1963) multi-cyclic Stationary Detection Delay (STADD) delivered by the SR procedure is with respect to the severity of the post-change distribution parameter misspecification. To answer this question, we offered a case study where the robustness of the SR procedure was examined numerically in a specific Gaussian scenario. The obtained exhaustive characterization of the SR procedure’s robustness can be used to develop the respective theory (which is still missing) and can also provide guidance to practitioners interested in employing the SR procedure. Qualitatively, the overall conclusion of the study was that the less (more) contrast the change and the lower (higher) the false alarm risk, the less (more) robust the SR procedure devised to detect it. This is an expected result.

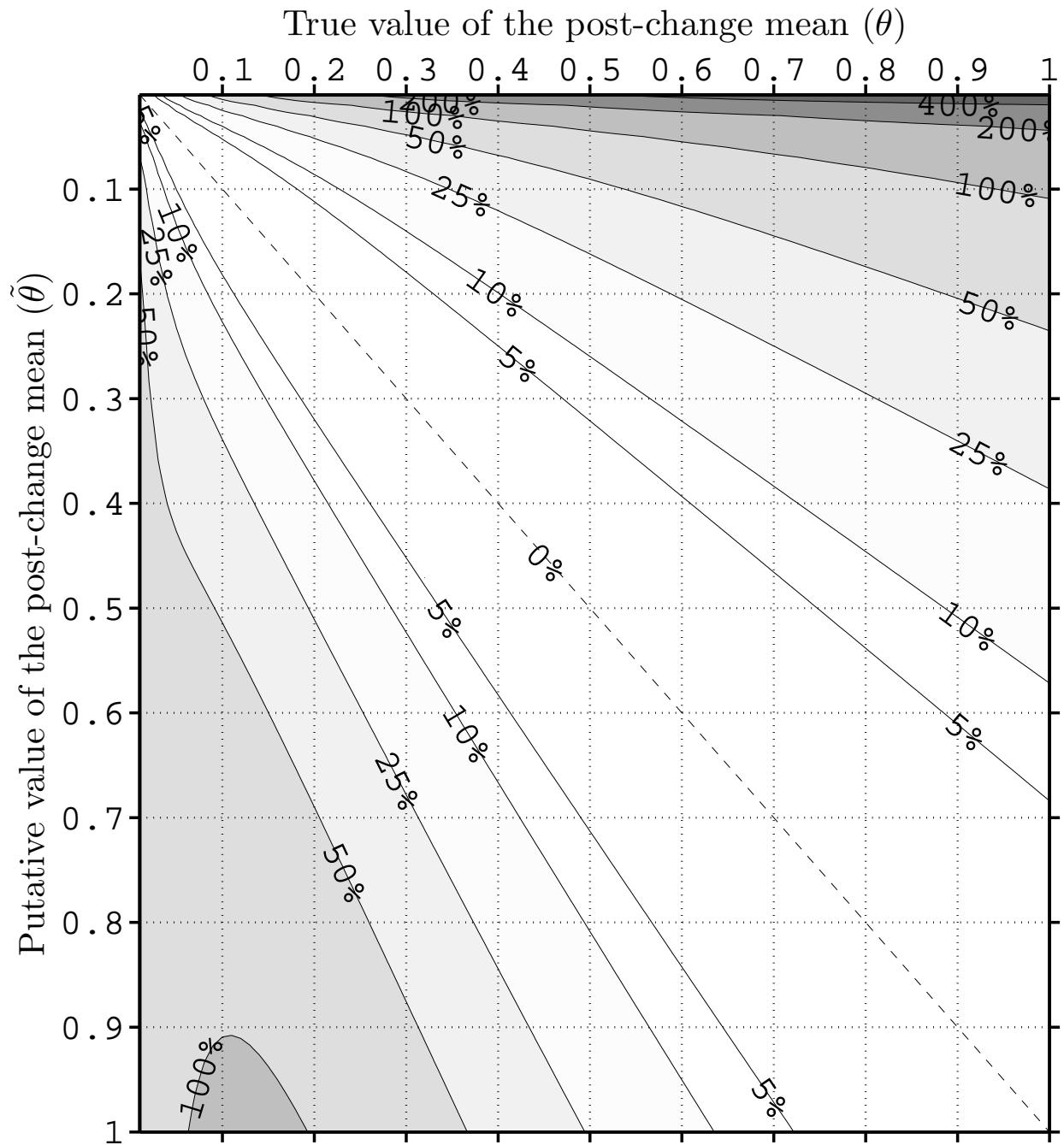


Figure 8: Characterization of $\text{RE}_{\tilde{\theta}, \theta}(S_A)$ as a function of the putative value ($\tilde{\theta}$) and of the true value (θ) of the post-change mean for the ARL to false alarm of 10^3 (i.e., $\text{ARL}(S_A^r) = \gamma = 10^3$).

Table 4: Characterization of STADD $_{\tilde{\theta},\theta}(\mathcal{S}_A)$ and RE $_{\tilde{\theta},\theta}(\mathcal{S}_A)$ as functions of the putative value ($\tilde{\theta}$) and of the true value (θ) of the post-change mean for the ARL to false alarm of 10^4 (i.e., $\text{ARL}(\mathcal{S}_A^r) = \gamma = 10^4$).

		True value of the post-change mean (θ)									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.1	516.46 (0 %)	214.84 (12.39 %)	134.68 (31.81 %)	98.04 (51.88 %)	77.09 (71.7 %)	63.53 (91.05 %)	54.04 (109.91 %)	47.02 (128.3 %)	41.63 (146.23 %)	37.36 (163.74 %)	
0.2	606.94 (17.52 %)	191.15 (0 %)	107.91 (5.61 %)	74.85 (15.95 %)	57.27 (27.55 %)	46.38 (39.49 %)	38.99 (51.45 %)	33.64 (63.3 %)	29.58 (74.98 %)	26.41 (86.48 %)	
0.3	804.97 (55.86 %)	205.88 (7.7 %)	102.18 (0 %)	66.66 (3.27 %)	49.33 (9.88 %)	39.14 (17.72 %)	32.45 (26.05 %)	27.72 (34.58 %)	24.2 (43.15 %)	21.49 (51.7 %)	
0.4	1,054.59 (104.2 %)	244.69 (28.01 %)	106.55 (4.28 %)	64.55 (0 %)	45.87 (2.16 %)	35.51 (6.79 %)	28.97 (12.53 %)	24.47 (18.78 %)	21.19 (25.3 %)	18.69 (31.94 %)	
0.5	1,326.24 (156.79 %)	303.4 (58.72 %)	118.7 (16.17 %)	66.3 (2.72 %)	44.9 (0 %)	33.76 (1.54 %)	27.03 (4.99 %)	22.53 (9.39 %)	19.33 (14.31 %)	16.93 (19.51 %)	
0.6	1,603.21 (210.42 %)	378.96 (98.25 %)	138.11 (35.17 %)	71.3 (10.46 %)	45.74 (1.88 %)	33.25 (0 %)	26.04 (1.15 %)	21.39 (3.83 %)	18.15 (7.33 %)	15.77 (11.32 %)	
0.7	1,876.54 (263.35 %)	468.46 (145.07 %)	164.59 (61.08 %)	79.43 (23.05 %)	48.18 (7.31 %)	33.71 (1.38 %)	25.74 (0 %)	20.78 (0.9 %)	17.42 (3.04 %)	15.0 (5.9 %)	
0.8	2,141.43 (314.64 %)	569.14 (197.74 %)	197.93 (93.71 %)	90.76 (40.61 %)	52.17 (16.21 %)	35.05 (5.4 %)	26.02 (1.06 %)	20.6 (0 %)	17.03 (0.72 %)	14.51 (2.47 %)	
0.9	2,394.87 (363.71 %)	678.44 (254.92 %)	237.81 (132.74 %)	105.4 (63.29 %)	57.79 (28.71 %)	37.24 (12.0 %)	26.81 (4.15 %)	20.77 (0.84 %)	16.91 (0 %)	14.25 (0.59 %)	
1.0	2,634.79 (410.16 %)	794.01 (315.38 %)	283.75 (177.7 %)	123.41 (91.2 %)	65.11 (45.02 %)	40.34 (21.31 %)	28.12 (9.25 %)	21.28 (3.29 %)	17.02 (0.68 %)	14.16 (0 %)	

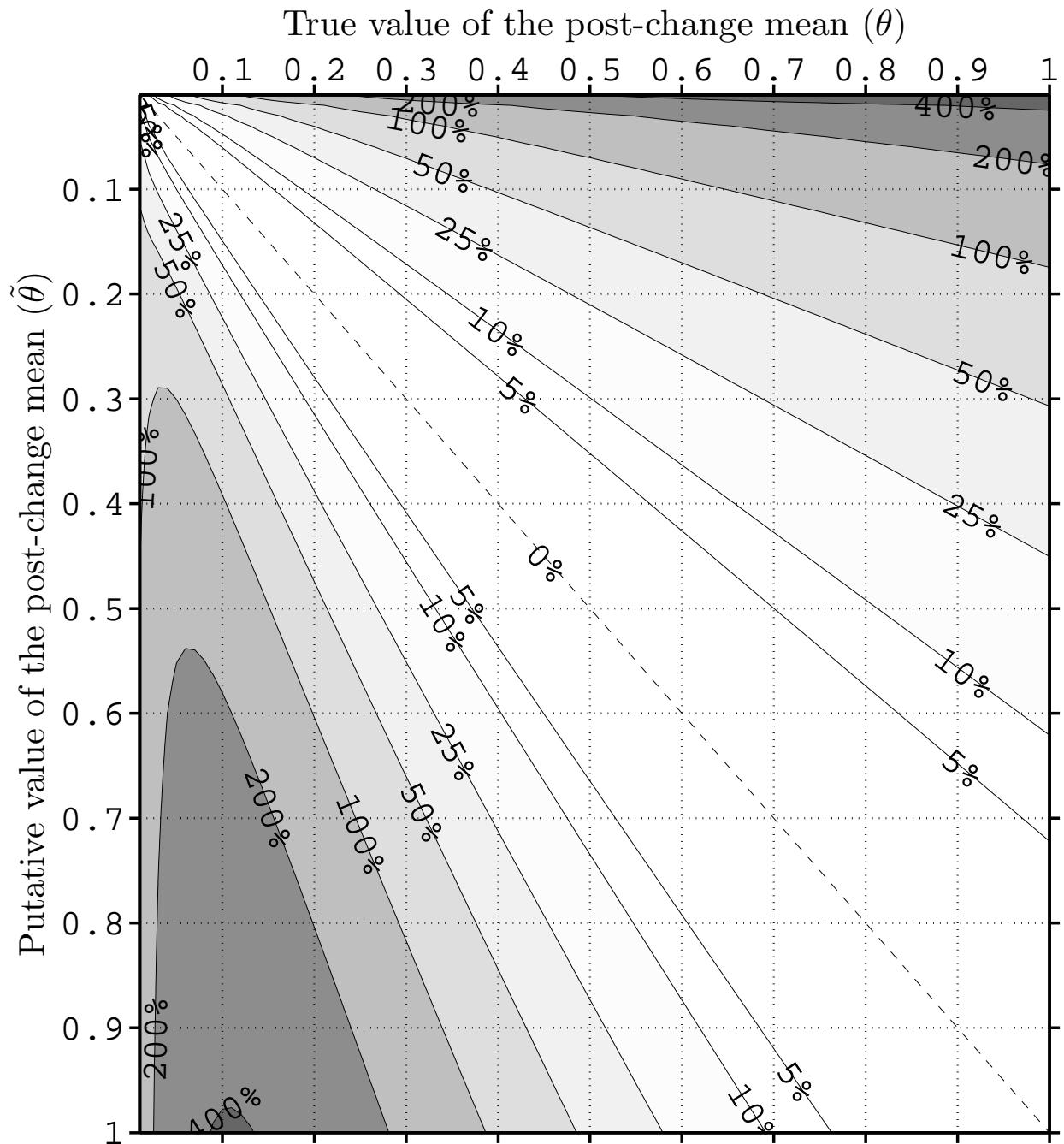


Figure 9: Characterization of $\text{RE}_{\tilde{\theta}, \theta}(S_A)$ as a function of the putative value ($\tilde{\theta}$) and of the true value (θ) of the post-change mean for the ARL to false alarm of 10^4 (i.e., $\text{ARL}(S_A^r) = \gamma = 10^4$).

Acknowledgements

The authors are grateful to Dr. Emmanuel Yashchin of the Mathematical Sciences Department at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, USA; to Prof. William H. Woodall of the Statistics Department at the Virginia Polytechnic Institute (Virginia Tech), Blacksburg, Virginia, USA; to Prof. Sven Knoth of the Department of Mathematics and Statistics at the Helmut Schmidt University, Hamburg, Germany; to Dean Neubauer of Corning Incorporated, Corning, New York, USA; to Prof. Subha Chakraborti of the Department of Information Systems, Statistics and Management Science at the University of Alabama, Alabama, USA; and to Dr. Ron Kenett of Israel-based KPA Ltd. (www.kpa-group.com), for the interest in this work and for the constructive feedback that helped improve the quality of the manuscript. Additional thanks goes out to the anonymous referee for the comments and suggestions that helped to ameliorate the paper further.

The effort of A.S. Polunchenko was supported, in part, by the Simons Foundation (www.simonsfoundation.org) via a Collaboration Grant in Mathematics (Award # 304574) and by the Research Foundation for the State University of New York at Binghamton via an Interdisciplinary Collaboration Grant (Award # 66761).

Last but not least, A.S. Polunchenko is also personally thankful to the Office of the Dean of the Harpur College of Arts and Sciences at the State University of New York (SUNY) at Binghamton for the support provided through the Dean's Research Semester Award for Junior Faculty granted for the Fall semester of 2014. The Award allowed to focus on this research more fully.

References

M. Basseville and I. V. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, Inc., Englewood Cliffs, NJ, April 1993. ISBN 0-13-126780-9.

A. J. Duncan. The economic design of \bar{X} charts used to maintain current control of a process. *Journal of the American Statistical Association*, 51(274):228–242, June 1956. doi: 10.1080/01621459.1956.10501322.

E. A. Feinberg and A. N. Shiryaev. Quickest detection of drift change for Brownian motion in generalized Bayesian and minimax settings. *Statistics & Decisions*, 24(4):445–470, 2006. doi: 10.1524/stnd.2006.24.4.445.

C. Ho and K. Case. Economic design of control charts: A literature review for 1981–1991. *Journal of Quality Technology*, 26(1):39–53, 1994.

R. S. Kenett and M. Pollak. Data-analytic aspects of the Shiryaev–Roberts control chart: Surveillance of a non-homogeneous Poisson process. *Journal of Applied Statistics*, 23(1):125–138, 1996. doi: 10.1080/02664769624413.

S. Knoth. The art of evaluating monitoring schemes—How to measure the performance of control charts? In H.-J. Lenz and P.-T. Wilrich, editors, *Frontiers in Statistical Quality Control 8*, pages 74–99. Physica-Verlag HD, 2006. doi: 10.1007/3-7908-1687-6_5.

G. Lorden. Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, 42(6):1897–1908, 1971. doi: 10.1214/aoms/1177693055.

T. J. Lorenzen and L. C. Vance. The economic design of control charts: A unified approach. *Technometrics*, 28(1):3–10, February 1986.

M. A. Mahmoud, W. H. Woodall, and R. E. Davis. Performance comparison of some likelihood ratio-based statistical surveillance methods. *Journal of Applied Statistics*, 35(7):783–798, July 2008. doi: 10.1080/02664760802005878.

D. C. Montgomery. The economic design of control charts: A review and literature survey. *Journal of Quality Technology*, 12(2):75–87, April 1980.

G. V. Moustakides. Optimal stopping times for detecting changes in distributions. *Annals of Statistics*, 14(4):1379–1387, 1986.

G. V. Moustakides, A. S. Polunchenko, and A. G. Tartakovsky. Numerical comparison of CUSUM and Shiryaev–Roberts procedures for detecting changes in distributions. *Communications in Statistics—Theory and Methods*, 38(16):3225–3239, 2009. doi: 10.1080/03610920902947774.

G. V. Moustakides, A. S. Polunchenko, and A. G. Tartakovsky. A numerical approach to performance analysis of quickest change-point detection procedures. *Statistica Sinica*, 21(2):571–596, April 2011.

E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1):100–115, 1954. doi: 10.1093/biomet/41.1-2.100.

M. Pollak. Optimal detection of a change in distribution. *Annals of Statistics*, 13(1):206–227, 1985. doi: 10.1214/aos/1176346587.

M. Pollak. Average run lengths of an optimal method of detecting a change in distribution. *Annals of Statistics*, 15(2):749–779, 1987. doi: 10.1214/aos/1176350373.

M. Pollak. The Shiryaev–Roberts changepoint detection procedure in retrospect—Theory and practice. In *Proceedings of the 2nd International Workshop in Sequential Methodologies (IWSM’2009)*, University of Technology of Troyes, Troyes, France, June 2009.

M. Pollak and D. Siegmund. A diffusion process and its applications to detecting a change in the drift of Brownian motion. *Biometrika*, 72(2):267–280, 1985. doi: 10.1093/biomet/72.2.267.

M. Pollak and A. G. Tartakovsky. Exact optimality of the Shiryaev–Roberts procedure for detecting changes in distributions. In *Proceedings of the 2008 International Symposium on Information Theory and Its Applications (ISITA’2008)*, pages 1–6, Auckland, New Zealand, December 2008. doi: 10.1109/ISITA.2008.4895424.

M. Pollak and A. G. Tartakovsky. Optimality properties of the Shiryaev–Roberts procedure. *Statistica Sinica*, 19(4):1729–1739, 2009.

A. S. Polunchenko and A. G. Tartakovsky. State-of-the-art in sequential change-point detection. *Methodology and Computing in Applied Probability*, 14(3):649–684, 2012. doi: 10.1007/s11009-011-9256-5.

A. S. Polunchenko, A. G. Tartakovsky, and N. Mukhopadhyay. Nearly optimal change-point detection with an application to cybersecurity. *Sequential Analysis*, 31(3):409–435, 2012. doi: 10.1080/07474946.2012.694351.

A. S. Polunchenko, G. Sokolov, and W. Du. Quickest change-point detection: A bird’s eye view. In *Proceedings of the 2013 Joint Statistical Meetings (JSM’2013)*, Montréal, Québec, Canada, August 2013.

A. S. Polunchenko, G. Sokolov, and W. Du. Efficient performance evaluation of the Generalized Shiryaev–Roberts detection procedure in a multi-cyclic setup. *Applied Stochastic Models in Business and Industry*, 30(6):723–739, 2014a. doi: 10.1002/asmb.2026.

A. S. Polunchenko, G. Sokolov, and W. Du. An accurate method for determining the pre change Run Length distribution of the Generalized Shiryaev–Roberts detection procedure. *Sequential Analysis*, 33(1):112–134, 2014b. doi: 10.1080/07474946.2014.856642.

A. S. Polunchenko, G. Sokolov, and A. G. Tartakovsky. Optimal design and analysis of the Exponentially Weighted Moving Average chart for exponential data. *Sri Lankan Journal of Applied Statistics*, 5(4):57–80, December 2015. doi: 10.4038/sljastats.v5i4.7784. Special Issue: Modern Statistical Methodologies in the Cutting Edge of Science.

H. V. Poor and O. Hadjiliadis. *Quickest Detection*. Cambridge University Press, New York, NY, 2009.

Y. Ritov. Decision theoretic optimality of the CUSUM procedure. *Annals of Statistics*, 18(3):1464–1469, 1990. doi: 10.1214/aos/1176347761.

S. W. Roberts. Control chart tests based on geometric moving averages. *Technometrics*, 1(3):239–250, August 1959. doi: 10.1080/00401706.1959.10489860.

S. W. Roberts. A comparison of some control chart procedures. *Technometrics*, 8(3):411–430, August 1966.

A. N. Shiryaev. The problem of the most rapid detection of a disturbance in a stationary process. *Soviet Mathematics—Doklady*, 2:795–799, 1961. Translation from Dokl. Akad. Nauk SSSR 138:1039–1042, 1961.

A. N. Shiryaev. On optimum methods in quickest detection problems. *Theory of Probability and Its Applications*, 8(1):22–46, January 1963. doi: 10.1137/1108002.

A. N. Shiryaev. *Optimal Stopping Rules*. Springer-Verlag, New York, NY, 1978.

A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, NY, second edition, 1995.

A. N. Shiryaev. Quickest detection problems in the technical analysis of the financial data. In H. Geman, D. Madan, S. R. Pliska, and T. Vorst, editors, *Mathematical Finance—Bachelier Congress 2000*, Springer Finance, pages 487–521. Springer Berlin Heidelberg, 2002. doi: 10.1007/978-3-662-12429-1_22.

A. N. Shiryaev and P. Y. Zryumov. On the linear and nonlinear generalized Bayesian disorder problem (discrete time case). In F. Delbaen, M. Rásonyi, and C. Stricker, editors, *Optimality and Risk—Modern Trends in Mathematical Finance*, pages 227–236. Springer Berlin Heidelberg, 2010. doi: 10.1007/978-3-642-02608-9_12.

M. Srivastava and Y. Wu. Comparison of EWMA, CUSUM and Shirayev–Roberts procedures for detecting a shift in the mean. *Annals of Statistics*, 21(2):645–670, 1993. doi: 10.1214/aos/1176349142.

A. Tartakovsky, I. Nikiforov, and M. Basseville. *Sequential Analysis: Hypothesis Testing and Change-point Detection*, volume 166 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, 2014.

A. G. Tartakovsky and G. V. Moustakides. State-of-the-art in Bayesian changepoint detection. *Sequential Analysis*, 29(2):125–145, April 2010. doi: 10.1080/07474941003740997.

A. G. Tartakovsky, A. S. Polunchenko, and G. V. Moustakides. Design and comparison of Shiryaev–Roberts- and CUSUM-type change-point detection procedures. In *Proceedings of the 2nd International Workshop in Sequential Methodologies (IWSM’2009)*, University of Technology of Troyes, Troyes, France, June 2009.

A. G. Tartakovsky, A. S. Polunchenko, and G. Sokolov. Efficient computer network anomaly detection by changepoint detection methods. *IEEE Journal of Selected Topics in Signal Processing*, 7(1):4–11, February 2013. doi: 10.1109/JSTSP.2012.2233713.

V. V. Veeravalli and T. Banerjee. Quickest change detection. In R. Chellappa and S. Theodoridis, editors, *Academic Press Library in Signal Processing: Array and Statistical Signal Processing*, volume 3, pages 209–256. Academic Press, Oxford, UK, 2013.

M. Woodroffe. *Nonlinear Renewal Theory in Sequential Analysis*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1982. ISBN 0-89871-180-0.